

A Uniform-grid Discretization Algorithm for Stochastic Control with Risk Constraints

Yin-Lam Chow, Marco Pavone

Abstract—In this paper, we present a discretization algorithm for finite horizon risk constrained dynamic programming algorithm in [1]. Although in a theoretical standpoint, Bellman’s recursion provides a systematic way to find optimal value functions and generate optimal history dependent policies, there is a serious computational issue. Even if the state space and action space of this constrained stochastic optimal control problem are finite, the spaces of risk threshold and the feasible risk update are closed bounded subset of real numbers. This prohibits any direct applications of unconstrained finite state iterative methods in dynamic programming found in [2]. In order to approximate Bellman’s operator derived in [1], we discretize the continuous action spaces and formulate a finite space approximation for the exact dynamic programming algorithm. We will also prove that the approximation error bound of optimal value functions is bound linearly by the step size of discretization. Finally, details for implementations and possible modifications are discussed.

I. INTRODUCTION

Constrained stochastic optimal control problems naturally arise in decision-making problems where one has to consider multiple objectives. Instead of introducing an aggregate utility function that has to be optimized, one consider a setup where one cost function is to be minimized while keeping the other cost functions below some given bounds. Application domains are broad and include engineering, finance, and logistics. Within a constrained framework, the most common setup is, arguably, the optimization of a *risk-neutral expectation* criterion subject to a *risk-neutral* constraint [3], [1]. This model, however, is not suitable in scenarios where risk-aversion is a key feature of the problem setup. To introduce risk aversion, in [1] the authors studied stochastic optimal control problems with risk constraints, where risk is modeled according to *dynamic, time-consistent risk metrics* [4], [5]. These metrics have the desirable property of ensuring rational consistency of risk preferences across multiple periods [5]. (In contrast, traditional static risk metrics, such as conditional value at risk, can lead to potentially “inconsistent” behaviors, see [6] and references therein.) In particular, in [1], the authors developed a dynamic programming approach that allows to (formally) compute the optimal costs by value iteration via a *constrained* dynamic programming operator. The key idea is that due to the compositional structure of dynamic risk constraints, the optimization problem can be cast as a Markov decision problem (MDP) on an augmented state space where Markov policies are optimal (as opposed to the original problem) and Bellman’s recursion can be applied. Henceforth, we will refer to such augmented MDP as AMDP. However, even if both the state space and action spaces for the original optimization problem are assumed to be finite, the augmented state in AMDP contains state

variables that are *continuous* and lie in bounded subsets of the real numbers. Hence, apart from a few cases when an analytical solution is available, the problem must be solved numerically.

Accordingly, the objective of this paper is to develop a numerical method for the solution of stochastic optimal control problems with dynamic, time-consistent risk measures. The approach is to discretize the continuous states in AMDP. Numerical algorithms for the solution of continuous MDPs is indeed a fairly mature field. In [7], [8], [9], multi-grid state/action space discretization methods are developed with bounds available on how fine the discretization should be in order to achieve a desired accuracy. In [10], the grid for discretization is chosen via randomized sampling techniques and Monte Carlo methods. In [11], the value functions are approximated by a finite number of basis functions. Variable resolution grid sampling techniques have been proposed in [12], [13], [14]. However, in general, these results assume that the dynamic programming operator is *unconstrained*, i.e., actions and future states are only constrained to lie in their respective feasible sets. In contrast, the dynamic programming operator for AMDP *constrains* actions and future states in a more complicated fashion (see Section II for more details). This precludes the application of current approximation algorithms to the numerical solution of AMDP.

Our approach is to extend the uniform grid discretization approximation developed in [9]. This requires the development of novel Lipschitz bounds for constrained dynamic programming operators. We show that convergence is linear in the step size, which is the same convergence rate for discretization algorithms for unconstrained dynamic programming operators [9]. The importance of our result is fourfold. First, we provide a sound numerical method for the solution of AMDP. Second, our results provide the basis to develop more sophisticated approximation algorithms (e.g., variable grid size, reinforcement learning, etc.) for the solution of stochastic optimal control problems with dynamic, time-consistent risk constraints. Third, a particular type or dynamic, time-consistent “risk” constraint is, of course, the risk neutral expectation. Hence, our results provide as a particular case a numerical algorithm to solve the dynamic programming equations that arise in traditional constrained stochastic optimal control problems [3]. To the best of our knowledge, this is the first practical algorithm to solve such dynamic programming equations. Finally, the ideas and techniques introduced in the current paper could be useful for the development of approximation algorithms for other types of constrained dynamic programming operators.

The rest of the paper is structured as follows. In Section II we present background material for this paper, in particular about dynamic, time-consistent risk metrics and stochastic optimal control with dynamic risk constraints [1]. In Section III we present and theoretically study a uniform

Y.-L. Chow and M. Pavone are with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA. Email: {ychow, pavone}@stanford.edu.

grid approximation algorithm for the augmented MDP; in particular, we show that the error bound is linear to the discretization step size. In Section IV, we study by numerical simulations the performance of the proposed algorithm and discuss details of implementations using Branch and Bound techniques. Finally, in Section V, we draw our conclusions and offer directions for future work.

II. PRELIMINARIES

In this section we provide some background for the theory of dynamic, time-consistent risk metrics and stochastic optimal control with dynamic risk constraints, on which we will rely extensively later in the paper.

A. Notations

In this paper, given a real-valued function f , $\text{dom}(f)$ denotes its domain and $\text{epi} f$ denotes its epigraph (i.e., the set of points lying on or above its graph). Let ν and μ be two probability measures on the same measurable space, then $\nu \ll \mu$ denotes that ν is absolutely continuous with respect to μ (i.e., $\nu(E) = 0$ for every set E for which $\mu(E) = 0$).

B. Markov Decision Processes

A finite Markov Decision Process (MDP) is a four-tuple $(S, U, Q, U(\cdot))$, where S , the state space, is a finite set; U , the control space, is a finite set; for every $x \in S$, $U(x) \subseteq U$ is a nonempty set which represents the set of admissible controls when the system state is x ; and, finally, $Q(\cdot|x, u)$ (the transition probability) is a conditional probability on S given the set of admissible state-control pairs, i.e., the sets of pairs (x, u) where $x \in S$ and $u \in U(x)$.

Define the space H_k of admissible histories up to time k by $H_k = H_{k-1} \times S \times U$, for $k \geq 1$, and $H_0 = S$. A generic element $h_{0,k} \in H_k$ is of the form $h_{0,k} = (x_0, u_0, \dots, x_{k-1}, u_{k-1}, x_k)$. Let Π be the set of all deterministic policies with the property that at each time k the control is a function of $h_{0,k}$. In other words, $\Pi := \{\{\pi_0 : H_0 \rightarrow U, \pi_1 : H_1 \rightarrow U, \dots\} | \pi_k(h_{0,k}) \in U(x_k) \text{ for all } h_{0,k} \in H_k, k \geq 0\}$.

C. Dynamic, time-consistent, risk measures

Consider a probability space (Ω, \mathcal{F}, P) , a filtration $\mathcal{F}_1 \subset \mathcal{F}_2 \dots \subset \mathcal{F}_N \subset \mathcal{F}$, and an adapted sequence of random variables $Z_k, k \in \{0, \dots, N\}$. We assume that $\mathcal{F}_0 = \{\Omega, \emptyset\}$, i.e., Z_0 is deterministic. In this paper we interpret the variables Z_k as stage-wise costs. For each $k \in \{1, \dots, N\}$, define the spaces of random variables with finite p th order moment as $\mathcal{Z}_k := L_p(\Omega, \mathcal{F}_k, P)$, $p \in [1, \infty]$; also, let $\mathcal{Z}_{k,N} := \mathcal{Z}_k \times \dots \times \mathcal{Z}_N$.

Roughly speaking, a dynamic risk measure is said time consistent if it is such that when a Z cost sequence is deemed less risky than a W cost sequence from the perspective of a future time k , and both sequences yield identical costs from the current time l to the future time k , then the Z sequence is deemed as less risky at the current time l . It turns out that dynamic, time-consistent risk metrics can be constructed by ‘‘compounding’’ one-step conditional risk measures, which are defined as follows.

Definition II.1 (Coherent one-step conditional risk measures). A coherent one-step conditional risk measures is a

mapping $\rho_k : \mathcal{Z}_{k+1} \rightarrow \mathcal{Z}_k, k \in \{0, \dots, N\}$, with the following four properties:

- *Convexity*: $\rho_k(\lambda Z + (1 - \lambda)W) \leq \lambda \rho_k(Z) + (1 - \lambda)\rho_k(W)$, $\forall \lambda \in [0, 1]$ and $Z, W \in \mathcal{Z}_{k+1}$;
- *Monotonicity*: if $Z \leq W$ then $\rho_k(Z) \leq \rho_k(W)$, $\forall Z, W \in \mathcal{Z}_{k+1}$;
- *Translation invariance*: $\rho_k(Z + W) = Z + \rho_k(W)$, $\forall Z \in \mathcal{Z}_k$ and $W \in \mathcal{Z}_{k+1}$;
- *Positive homogeneity*: $\rho_k(\lambda Z) = \lambda \rho_k(Z)$, $\forall Z \in \mathcal{Z}_{k+1}$ and $\lambda \geq 0$.

Then, the following results characterize dynamic, time-consistent risk metrics [4].

Theorem II.2 (Dynamic, time-consistent risk measures). Consider, for each $k \in \{0, \dots, N\}$, the mappings $\rho_{k,N} : \mathcal{Z}_{k,N} \rightarrow \mathcal{Z}_k$ defined as

$$\rho_{k,N} = Z_k + \rho_k(Z_{k+1} + \rho_{k+1}(Z_{k+2} + \dots + \rho_{N-2}(Z_{N-1} + \rho_{N-1}(Z_N)) \dots)), \quad (1)$$

where the ρ_k 's are coherent one-step risk measures. Then, the ensemble of such mappings is a time-consistent dynamic risk measure.

In this paper we consider a (slight) refinement of the concept of dynamic, time-consistent risk measure, which involves the addition of a Markovian structure [4].

Definition II.3 (Markov dynamic risk measures). Let $\mathcal{V} := L_p(S, \mathcal{B}, P)$ be the space of random variables on S with finite p th moment. Given a controlled Markov process $\{x_k\}$, a Markov dynamic risk measure is a dynamic, time-consistent risk measure if each coherent one-step risk measure $\rho_k : \mathcal{Z}_{k+1} \rightarrow \mathcal{Z}_k$ in equation (1) can be written as:

$$\rho_k(V(x_{k+1})) = \sigma_k(V(x_{k+1}), x_k, Q(x_{k+1}|x_k, u_k)), \quad (2)$$

for all $V(x_{k+1}) \in \mathcal{V}$ and $u \in U(x_k)$, where σ_k is a coherent one-step risk measure on \mathcal{V} (with the additional technical property that for every $V(x_{k+1}) \in \mathcal{V}$ and $u \in U(x_k)$ the function $x_k \mapsto \sigma_k(V(x_{k+1}), x_k, Q(x_{k+1}|x_k, u_k))$ is an element of \mathcal{V}).

In other words, in a Markov dynamic risk measures, the evaluation of risk is not allowed to depend on the whole past.

D. Stochastic optimal control with dynamic, time-consistent risk constraints

Consider an MDP and let $c : S \times U \rightarrow \mathbb{R}$ and $d : S \times U \rightarrow \mathbb{R}$ be functions which denote costs associated with state-action pairs. Given a policy $\pi \in \Pi$, an initial state $x_0 \in S$, and an horizon $N \geq 1$, the cost function is defined as

$$J_N^\pi(x_0) := \mathbb{E} \left[\sum_{k=0}^{N-1} c(x_k, u_k) \right],$$

and the risk constraint is defined as

$$R_N^\pi(x_0) := \rho_{0,N}(d(x_0, u_0), \dots, d(x_{N-1}, u_{N-1}), 0),$$

where $\rho_{k,N}(\cdot), k \in \{0, \dots, N-1\}$, is a Markov dynamic risk measure (for simplicity, we do not consider terminal costs, even though their inclusion is straightforward). The problem is then as follows:

Optimization problem \mathcal{OPT} — Given an initial state $x_0 \in S$, a time horizon $N \geq 1$, and a risk threshold $r_0 \in \mathbb{R}$, solve

$$\begin{aligned} \min_{\pi \in \Pi} \quad & J_N^\pi(x_0) \\ \text{subject to} \quad & R_N^\pi(x_0) \leq r_0. \end{aligned}$$

If problem \mathcal{OPT} is not feasible, we say that its value is ∞ . In [1] the authors developed a dynamic programming approach to solve this problem. To define the value functions, one needs to define the tail subproblems. For a given $k \in \{0, \dots, N-1\}$ and a given state $x_k \in S$, we define the *sub-histories* as $h_{k,j} := (x_k, u_k, \dots, x_j)$ for $j \in \{k, \dots, N\}$; also, we define the *space of truncated policies* as $\Pi_k := \left\{ \{\pi_k, \pi_{k+1}, \dots\} \mid \pi_j(h_{k,j}) \in U(x_j) \text{ for } j \geq k \right\}$. For a given stage k and state x_k , the cost of the tail process associated with a policy $\pi \in \Pi_k$ is simply $J_N^\pi(x_k) := \mathbb{E} \left[\sum_{j=k}^{N-1} c(x_j, u_j) \right]$. The risk associated with the tail process is:

$$R_N^\pi(x_k) := \rho_{k,N} \left(d(x_k, u_k), \dots, d(x_{N-1}, u_{N-1}), 0 \right),$$

The tail subproblems are then defined as

$$\min_{\pi \in \Pi_k} J_N^\pi(x_k) \quad (3)$$

$$\text{subject to} \quad R_N^\pi(x_k) \leq r_k(x_k), \quad (4)$$

for a given (undetermined) threshold value $r_k(x_k) \in \mathbb{R}$ (i.e., the tail subproblems are specified up to a threshold value).

For each $k \in \{0, \dots, N-1\}$ and $x_k \in S$, we define the set of feasible constraint thresholds as

$$\Phi_k(x_k) := [\underline{R}_N(x_k), \overline{R}_{N,k}], \quad \Phi_N(x_N) := \{0\},$$

where $\underline{R}_N(x_k) := \min_{\pi \in \Pi_k} R_N^\pi(x_k)$, and $\overline{R}_{N,k} = (N-k)\rho_{\max}$. The value functions are then defined as follows:

- If $k < N$ and $r_k \in \Phi_k(x_k)$:

$$\begin{aligned} V_k(x_k, r_k) &= \min_{\pi \in \Pi_k} J_N^\pi(x_k) \\ \text{subject to} \quad & R_N^\pi(x_k) \leq r_k. \end{aligned}$$

- if $k \leq N$ and $r_k \notin \Phi_k(x_k)$:

$$V_k(x_k, r_k) = \infty;$$

- when $k = N$ and $r_N = 0$:

$$V_N(x_N, r_N) = 0.$$

Let $B(S)$ denote the space of real-valued bounded functions on S , and $B(S \times \mathbb{R})$ denote the space of real-valued bounded functions on $S \times \mathbb{R}$. For $k \in \{0, \dots, N-1\}$, we define the dynamic programming operator $T_k[V_{k+1}] : B(S \times \mathbb{R}) \mapsto B(S \times \mathbb{R})$ according to the equation:

$$\begin{aligned} T_k[V_{k+1}](x_k, r_k) &:= \inf_{(u, r') \in F_k(x_k, r_k)} \left\{ c(x_k, u) + \right. \\ &\quad \left. \sum_{x_{k+1} \in S} Q(x_{k+1} | x_k, u) V_{k+1}(x_{k+1}, r'(x_{k+1})) \right\}, \end{aligned} \quad (5)$$

where F_k is the set of control/threshold *functions*:

$$F_k(x_k, r_k) := \left\{ (u, r') \mid \begin{aligned} &u \in U(x_k), r'(x') \in \Phi_{k+1}(x') \text{ for} \\ &\text{all } x' \in S, \text{ and } d(x_k, u) + \rho_k(r'(x_{k+1})) \leq r_k \end{aligned} \right\}.$$

If $F_k(x_k, r_k) = \emptyset$, then $T_k[V_{k+1}](x_k, r_k) = \infty$.

Note that, for a given state and threshold constraint, set F_k characterizes the set of feasible pairs of actions and subsequent constraint thresholds.

Theorem II.4 (Bellman's equation with risk constraints). *For all $k \in \{0, \dots, N-1\}$ the optimal cost functions satisfy the Bellman's equation:*

$$V_k(x_k, r_k) = T_k[V_{k+1}](x_k, r_k).$$

E. Representation theorems

A key result that will be heavily exploited in this paper is the following representation theorem for coherent one-step conditional risk measures.

Theorem II.5. $\rho_k : \mathcal{Z}_{k+1} \rightarrow \mathcal{Z}_k$ is a coherent one-step conditional risk measure if and only if

$$\rho_k(Z(x_{k+1})) = \sup_{\xi \in \mathcal{U}_{k+1}(x_k, Q(x_{k+1} | x_k, u_k))} \sum_{x' \in S} \xi(x') Z(x') \quad (6)$$

where

$$\mathcal{U}_{k+1}(x_k, Q) =$$

$$\left\{ \xi \in \mathcal{M} \mid \xi \ll Q, \sum_{x' \in S} \xi(x') Z(x') \leq \rho_k(Z), \forall Z \in \mathcal{Z}_{k+1} \right\},$$

and

$$\mathcal{M} = \left\{ \xi \in \mathbb{R}^{|S|} \mid \sum_{x' \in S} \xi(x') = 1, \xi(x') \geq 0, \forall x' \in S \right\}.$$

For $Z \in \mathcal{Z}_k \subseteq L_p(\Omega, \mathcal{F}, P)$, $\mathcal{U}_{k+1}(x_k, Q)$ is a subset of $L_q(\Omega, \mathcal{F}, P)$, where $L_q(\Omega, \mathcal{F}, P)$ is the dual space of $L_p(\Omega, \mathcal{F}, P)$, for $1/p + 1/q = 1$ and $p, q \in [1, \infty]$.

Proof. Refer to Theorem 6.4 in [5] and references therein. \square

The result essentially says that any coherent risk measure can be interpreted as an expectation taken with respect to a worst-case measure, which is chosen from a suitable set of test measures [6].

Furthermore, by Moreau-Rockafellar Theorem (Theorem 7.4 in [5]), it implies $\mathcal{U}_{k+1}(x_k, Q(x_{k+1} | x_k, u_k)) = \partial \rho_k(0)$, when the transition probability kernel is $Q(x_{k+1} | x_k, u_k)$. The next Theorem implies a basic duality result on coherent risk measures.

Theorem II.6. $\rho_k : \mathcal{Z}_{k+1} \rightarrow \mathcal{Z}_k$ is a coherent one-step conditional risk measure, if and only if there exists a bounded, non-empty, weakly* compact and convex set: $\mathcal{U}_{k+1}(x_k, Q(x_{k+1} | x_k, u_k))$ such that equation (6) holds. Furthermore, if ρ_k is a coherent risk measure, then it is continuous and sub-differentiable in \mathcal{Z}_{k+1} , also if $\text{dom}(\rho_k) = \{Z \in \mathcal{Z}_{k+1} : \rho_k(Z) < \infty\}$ has an non-empty interior, then ρ_k is finite valued.

Proof. See Proposition 6.5, Theorem 6.6 and Theorem 6.7 in [5]. \square

Since the analysis of this paper is restricted to finite state and action spaces, from this theorem, $\mathcal{U}_{k+1}(x_k, Q(x_{k+1}|x_k, u_k)) = \partial \rho_k(0)$ is a non-empty, convex, bounded and compact set in $\mathbb{R}^{|S|}$. By extreme value theorem, the supremum in equation (6) is attained.

III. DISCRETIZATION OF THE CONTINUOUS RISK THRESHOLDS IN CONSTRAINED DYNAMIC PROGRAMMING

In the previous section, we have shown that the constrained stochastic optimal control problem can be solved using value iteration (See Theorem II.4). However, the constant risk threshold r_k in value function $V_k(x_k, r_k)$, $k \in \{0, \dots, N-1\}$ is a continuous state. This results in numerical complexity when value iteration is performed. Therefore, in this section, we consider a numerical approximation algorithm using discretization. First of all, from the dynamic programming operator possess several nice properties:

Lemma III.1. *Let $V, \tilde{V} \in B(S \times \mathbb{R})$ be real-valued bounded functions and $T_k[V] : B(S \times \mathbb{R}) \mapsto B(S \times \mathbb{R})$ be a dynamic programming operator in given in $B(S \times \mathbb{R})$ whose expression is given by equation (5) for any $k \in \{0, \dots, N-1\}$. Then, the following statement holds:*

- 1) **Monotonicity:** *For any $(x, r) \in B(S \times \mathbb{R})$, if $V \leq \tilde{V}$, then $T_k[V](x, r) \leq T_k[\tilde{V}](x, r)$.*
- 2) **Constant shift:** *For any real number L and $(x, r) \in B(S \times \mathbb{R})$, $T_k[V + L](x, r) = T_k[V](x, r) + L$, where $(V + L)(x, r) := V(x, r) + L$, $\forall (x, r) \in B(S \times \mathbb{R})$.*
- 3) **Non-expansivity:** *For all $V, \tilde{V} \in B(S \times \mathbb{R})$, $\|T_k[V] - T_k[\tilde{V}]\|_\infty \leq \|V - \tilde{V}\|_\infty$, where $\|\cdot\|_\infty$ is the infinity norm of a function.*

Next, we introduce the method for constrained dynamic programming with discretized risk thresholds and updates.

A. Dynamic programming with discretized risk thresholds and updates

For $k \in \{0, \dots, N-1\}$, we will partition $\Phi_k(x_k)$ into $t+1$ partitions using t grid points: $\{\hat{r}_k^{(1)}, \dots, \hat{r}_k^{(t)}\}$ for every fixed $x_k \in S$. The step size of discretization of the risk thresholds r_k is Δ . For $\tau \in \{0, \dots, t\}$, define the discretized region $\Phi_k^{(\tau)}(x_k) = [r_k^{(\tau)}, r_k^{(\tau+1)})$, where $r_k^{(0)} = \underline{R}_N(x_k)$ and $r_k^{(t+1)} = \bar{R}_{N,k} + \epsilon$, for arbitrarily small $\epsilon > 0$. We also define $\bar{\Phi}_k(x_k) = \{r_k^{(0)}, \dots, r_k^{(t+1)}\}$ to be a finite state of risk threshold at step k . Let $\tau \in \{0, \dots, t\}$ such that $r_k \in \Phi_k^{(\tau)}(x_k)$. Now, define the approximation operator $T_{\Delta,k}^{D,\tau}$ for $x_k \in S$, $r_k \in \Phi_k^{(\tau)}(x_k)$:

$$T_{\Delta,k}^D[V](x_k, r_k) := \bar{T}_{\Delta,k}^D[V](x_k, r_k^{(\tau)}) \quad (7)$$

where

$$\begin{aligned} \bar{T}_{\Delta,k}^D[V](x_k, r_k) := & \min_{(u, r^{D,\tau}) \in F_k^D(x_k, r_k)} \left\{ c(x_k, u) \right. \\ & \left. + \sum_{x' \in S} Q(x'|x_k, u) V(x', r^{D,\tau}(x')) \right\}, \end{aligned} \quad (8)$$

where F_k^D is the set of control/threshold functions:

$$F_k^D(x_k, r_k) := \left\{ (u, r') \mid u \in U(x_k), r^{D,\tau}(x') \in \bar{\Phi}_{k+1}(x'), \right. \\ \left. \forall x' \in S, d(x_k, u) + \rho_k(r^{D,\tau}(x_{k+1})) \leq r_k \right\}.$$

If $F_k^D(x_k, r_k) = \emptyset$, then $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k) = \infty$.

By construction, we can see the set of optimal solution of $T_{\Delta,k}^D[V](x_k, r_k)$ is a subset of feasible space for the problem described by $T_k[V](x_k, r_k)$ (since $F_k^D(x_k, r_k) \subseteq F_k(x_k, r_k)$ and $r_k^{(\tau)} \leq r_k$). Because the solution of $T_{\Delta,k}^D[V](x_k, r_k)$ is an infimum over a finite set, the problem in (7) is a minimization. Also, based on similar proofs, the dynamic programming operator $T_{\Delta,k}^D[V]$ satisfies all the properties given in Lemma III.1. The main result of this section is to obtain a bound of the differences between $T_k[V](x_k, r_k)$ and $T_{\Delta,k}^D[V](x_k, r_k)$, which will be given in the next subsection.

B. Error bound analysis

First, we have the following assumptions for the following analysis:

Assumptions for discretization analysis:

- 1) There exists $M_c, M_d > 0$ such that

$$|c(x, u) - c(x, \tilde{u})| \leq M_c |u - \tilde{u}|,$$

$$|d(x, u) - d(x, \tilde{u})| \leq M_d |u - \tilde{u}|,$$

for any $x \in S$, $u, \tilde{u} \in U(x)$.

- 2) For any $u, \tilde{u} \in U(x_k)$, there exists $M_q > 0$ such that

$$\sum_{x' \in S} |Q(x'|x_k, u) - Q(x'|x_k, \tilde{u})| \leq M_q |u - \tilde{u}|.$$

Assumptions (1) to (2) are the critical assumptions required to perform error bound analysis in this section. First, we have following Proposition showing the Lipschitz-ness of set-valued mapping $\mathcal{U}_{k+1}(x_k, Q)$.

Proposition III.2. *For any $\xi \in \mathcal{U}_{k+1}(x_k, Q)$, there exists a $M_\xi > 0$ such that for some $\tilde{\xi} \in \mathcal{U}_{k+1}(x_k, \tilde{Q})$,*

$$\sum_{x' \in S} |\xi(x') - \tilde{\xi}(x')| \leq M_\xi \sum_{x' \in S} |Q(x') - \tilde{Q}(x')|.$$

Proof. From Theorem II.6, we know that $\mathcal{U}_{k+1}(x_k, Q)$ is a closed, bounded, convex set of probability mass functions. Since any conditional probability mass function Q is in the interior of $\text{dom}(\mathcal{U}_{k+1})$ and the graph of $\mathcal{U}_{k+1}(x_k, Q)$ is closed, by Theorem 2.7 in [15], $\mathcal{U}_{k+1}(x_k, Q)$ is a Lipschitz set-valued mapping with respect to the Hausdorff distance. Thus, for any $\xi \in \mathcal{U}_{k+1}(x_k, Q)$, the following expression holds for some $M_\xi > 0$:

$$\inf_{\tilde{\xi} \in \mathcal{U}_{k+1}(x_k, \tilde{Q})} \sum_{x' \in S} |\xi(x') - \tilde{\xi}(x')| \leq M_\xi \sum_{x' \in S} |Q(x') - \tilde{Q}(x')|.$$

Next, we want to show that the infimum of the left side is attained. Since the objective function is convex, and $\mathcal{U}_{k+1}(x_k, \tilde{Q})$ is a convex compact set, there exists $\tilde{\xi} \in \mathcal{U}_{k+1}(x_k, \tilde{Q})$ such that infimum is attained. \square

Next, we provide a Lemma that characterizes an upper bound for the magnitude of the value functions.

Lemma III.3. For $k \in \{0, \dots, N-1\}$, the following bound is given for the value function $V_k(x_k, r_k)$: $\|V_k\|_\infty \leq (N - k)c_{\max}$, where

$$c_{\max} := \max_{(x,u) \in S \times U} |c(x, u)|. \quad (9)$$

Proof. First, from the definition of $V_N(x_N, r_N)$, we know that $V_N(x_N, r_N) = 0$ for any $x_N \in S$, $r_N \in \Phi_N(x_N)$. Therefore, the above inequality holds for the for $k = N$. For $j \in \{0, \dots, N-1\}$, since $|c(x_j, u_j)| \leq c_{\max}$ for any $x_j \in S$, $u_j \in U(x_j)$, it implies $\|T_j[V_N]\|_\infty \leq c_{\max}$. Furthermore,

$$\begin{aligned} \|V_j\|_\infty &= \|V_j - V_N\|_\infty \\ &\leq \|T_j[V_{j+1}] - T_j[V_N]\|_\infty + \|T_j[V_N] - V_N\|_\infty \\ &\leq \|V_{j+1} - V_N\|_\infty + c_{\max} = \|V_{j+1}\|_\infty + c_{\max}. \end{aligned}$$

The first inequality is due to triangle inequality and Theorem II.4, the second inequality is due to the non-expansivity property in Lemma III.1 and both equalities in the above expression are due to $V_N(x, r) = 0$. Thus by recursion, we get

$$\|V_k\|_\infty = \sum_{j=k}^{N-1} (\|V_j\|_\infty - \|V_{j+1}\|_\infty).$$

and the proof is completed by noting that $\|V_j\|_\infty - \|V_{j+1}\|_\infty \leq c_{\max}$ for $j \in \{k, \dots, N-1\}$. \square

To prove the main result, we need the following technical Lemma.

Lemma III.4. For every given $x_k \in S$ and $\tilde{r}_k, r_k \in \Phi_k(x_k)$, suppose Assumptions (1) to (2) hold. Also, define $\underline{r}' := \{r'(x')\}_{x' \in S} \in \mathbb{R}^{|S|}$ and $\underline{\hat{r}}' := \{\hat{r}'(x')\}_{x' \in S} \in \mathbb{R}^{|S|}$. If $F_k(x_k, r_k)$ and $F_k(x_k, \tilde{r}_k)$ are non-empty sets, then for any $(u, \underline{r}') \in F_k(x_k, r_k)$, there exists $(\hat{u}, \underline{\hat{r}}') \in F_k(x_k, \tilde{r}_k)$ such that for some $M_{r,k} > 0$,

$$|u - \hat{u}| + \sum_{x' \in S} |r'(x') - \hat{r}'(x')| \leq M_{r,k} |r_k - \tilde{r}_k|. \quad (10)$$

Proof. First, we want to show that $\alpha(u, \underline{r}') := d(x_k, u) + \rho_k(r'(x_{k+1}))$ is a Lipschitz function. Define $\{\xi^*(x')\}_{x' \in S} \in \arg \max_{\xi \in \mathcal{U}_{k+1}(x_k, Q(x_{k+1}|x_k, u))} \left\{ d(x_k, u) + \sum_{x' \in S} \xi(x')(r'(x')) \right\}$. Then, there exists a $\tilde{\xi} \in \mathcal{U}_{k+1}(x_k, Q(x_{k+1}|x_k, \tilde{u}))$ such that the following expressions hold:

$$\begin{aligned} &\alpha(u, \underline{r}') - \alpha(\tilde{u}, \underline{\tilde{r}}') \\ &= d(x_k, u) + \rho_k(r'(x_{k+1})) - d(x_k, \tilde{u}) - \rho_k(\tilde{r}'(x_{k+1})) \\ &\leq d(x_k, u) - d(x_k, \tilde{u}) + \sum_{x' \in S} (\xi^*(x') - \tilde{\xi}(x')) r'(x') \\ &\quad + \sum_{x' \in S} \tilde{\xi}(x') (r'(x') - \tilde{r}'(x')) \\ &\leq |d(x_k, u) - d(x_k, \tilde{u})| + \sum_{x' \in S} |r'(x') - \tilde{r}'(x')| \\ &\quad + \max_{x \in S} |r'(x)| \sum_{x' \in S} |\xi^*(x') - \tilde{\xi}(x')|. \end{aligned} \quad (11)$$

The first equality follows from definitions of coherent risk measures. The first inequality is due to Theorem II.5 and

the definition of $\xi^* \in \mathcal{U}_{k+1}(x_k, Q(x_{k+1}|x_k, u))$. The second inequality is due to the fact that ξ is a probability mass functions in $\mathcal{U}_{k+1}(x_k, Q(x_{k+1}|x_k, \tilde{u}))$. Then, by Proposition III.2, there exists $M_\xi > 0$ such that

$$\sum_{x' \in S} |\xi^*(x') - \tilde{\xi}(x')| \leq M_\xi \sum_{x' \in S} |Q(x'|x_k, u) - Q(x'|x_k, \tilde{u})|.$$

Furthermore, by Assumptions (1) to (2) and the definition of $\Phi_{k+1}(x_{k+1})$, expression (11) implies

$$\alpha(u, \underline{r}') - \alpha(\tilde{u}, \underline{\tilde{r}}') \leq M_{A,k} \left(|\tilde{u} - u| + \sum_{x' \in S} |\tilde{r}'(x') - r'(x')| \right)$$

where

$$M_{A,k} = \max \{M_d + M_q(\bar{R}_{N,k})M_\xi, 1\}.$$

By a symmetric argument, we can also show that

$$\alpha(\tilde{u}, \underline{\tilde{r}}') - \alpha(u, \underline{r}') \leq M_{A,k} \left(|\tilde{u} - u| + \sum_{x' \in S} |\tilde{r}'(x') - r'(x')| \right).$$

Thus, by combining both arguments, we have shown that $\alpha(u, \underline{r}')$ is a Lipschitz function. Next, for any $(u, r') \in F_k(x_k, r_k)$, where

$$F_k(x_k, r) = \left\{ (u, r') \mid u \in U(x_k), r'(x') \in \Phi_{k+1}(x'), \forall x' \in S, \alpha(u, \underline{r}') \leq r \right\},$$

consider the following optimization problem:

$$P_{x_k, u, \underline{r}'}(r) = \inf_{(\tilde{u}, \underline{\tilde{r}}') \in F_k(x_k, r)} |\tilde{u} - u| + \sum_{x' \in S} |\tilde{r}'(x') - r'(x')|.$$

Since (u, \underline{r}') is a feasible point of $F_k(x_k, r_k)$, $P_{x_k, u, \underline{r}'}(r_k) = 0$. By our assumptions, both $U(x_k)$ and $\Phi_{k+1}(x_{k+1})$ are compact sets of real numbers. Note that both $|\tilde{u} - u| + \sum_{x' \in S} |\tilde{r}'(x') - r'(x')|$ and $\alpha(\tilde{u}, \underline{\tilde{r}}')$ are Lipschitz functions in $(\tilde{u}, \underline{\tilde{r}}')$. Also, consider the sub-gradient of $f(\tilde{u}, \underline{\tilde{r}}', r) := \alpha(\tilde{u}, \underline{\tilde{r}}') - r$:

$$\begin{aligned} \partial f(\tilde{u}, \underline{\tilde{r}}', r) &= \bigcap_{(\hat{u}, \underline{\hat{r}}', \hat{r}) \in \text{dom}(f)} \left\{ \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \in \mathbb{R}^{|U|} \times \mathbb{R}^{|S|} \times \mathbb{R} : \right. \\ &\quad \left. f(\hat{u}, \underline{\hat{r}}', \hat{r}) \geq f(\tilde{u}, \underline{\tilde{r}}', r) + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}^T \left(\begin{bmatrix} \hat{u} \\ \underline{\hat{r}}' \\ \hat{r} \end{bmatrix} - \begin{bmatrix} \tilde{u} \\ \underline{\tilde{r}}' \\ r \end{bmatrix} \right) \right\}. \end{aligned}$$

Since $\alpha(\tilde{u}, \underline{\tilde{r}}') - r$ is differentiable on r , the third element of $\partial f(\tilde{u}, \underline{\tilde{r}}', r)$ is a singleton and it equals to $\{-1\}$. Next, consider the sub-gradient of $h(\tilde{u}, \underline{\tilde{r}}', r) = |\tilde{u} - u| + \sum_{x' \in S} |\tilde{r}'(x') - r'(x')|$. By identical arguments, we can show that the set of the third element of $\partial h(\tilde{u}, \underline{\tilde{r}}', r)$ is a singleton and it equals to $\{0\}$. Therefore, Theorem 4.2 in [16] implies $P_{x_k, u, \underline{r}'}(r)$ is strictly differentiable (Lipschitz continuous) in

¹ A sub-gradient of a function $f: \mathcal{X} \rightarrow \mathbb{R}$ at a point $x_0 \in \mathcal{X}$ is a real vector g such that for all $x \in \mathcal{X}$, $f(x) - f(x_0) \geq g^T(x - x_0)$, $\forall x \in \mathcal{X}$.

r^2 . Then, for any $(u, \underline{r}') \in F_k(x_k, r_k)$, there exists $M_{r,k} > 0$ such that

$$\inf_{(\tilde{u}, \tilde{\underline{r}}') \in F_k(x_k, \tilde{r}_k)} |\tilde{u} - u| + \sum_{x' \in S} |r'(x') - \tilde{r}'(x')| \leq M_{r,k} |\tilde{r}_k - r_k|.$$

Finally we want to show that the infimum on the left side of the above expression is attained. First, $|\tilde{u} - u| + \sum_{x' \in S} |r'(x') - \tilde{r}'(x')|$ is coercive and continuous in $(\tilde{u}, \tilde{\underline{r}}')$. By Example 14.29 in [17], this function is a Caratheodory integrand and is also a normal integrand. Furthermore, since $F_k(x_k, \tilde{r}_k)$ is a closed set (since $U(x_k)$ is a finite set, $\Phi_{k+1}(x_{k+1})$ is a compact set and the constraint inequality is non-strict) and $\alpha(\tilde{u}, \tilde{\underline{r}}') - \tilde{r}_k$ is a normal integrand (see the proof of Theorem IV.2 in [1]), by Theorem 14.36 and Example 14.32 in [17], one can show that the following indicator function:

$$\mathbb{I}_{x_k}(\tilde{u}, \tilde{\underline{r}}', \tilde{r}_k) := \begin{cases} 0 & \text{if } (\tilde{u}, \tilde{\underline{r}}') \in F_k(x_k, \tilde{r}_k) \\ \infty & \text{otherwise} \end{cases}$$

is a normal integrand. Furthermore, By Proposition 14.44 in [17], the function

$$g_{x_k}(\tilde{u}, \tilde{\underline{r}}', \tilde{r}_k) := |\tilde{u} - u| + \sum_{x' \in S} |r'(x') - \tilde{r}'(x')| + \mathbb{I}_{x_k}(\tilde{u}, \tilde{\underline{r}}', \tilde{r}_k)$$

is a normal integrand. Also, $\inf_{\tilde{u}} g_{x_k}(\tilde{u}, \tilde{\underline{r}}', \tilde{r}_k) = \inf_{(\tilde{u}, \tilde{\underline{r}}') \in F_k(x_k, \tilde{r}_k)} |\tilde{u} - u| + \sum_{x' \in S} |r'(x') - \tilde{r}'(x')|$. By Theorem 14.37 in [17], there exists $(\hat{u}, \hat{\underline{r}}') \in F_k(x_k, \tilde{r}_k)$ such that $(\hat{u}, \hat{\underline{r}}')$ argmin $g_{x_k}(\tilde{u}, \tilde{\underline{r}}', \tilde{r}_k)$. Furthermore, the right side of the above equality is finite since $F_k(x_k, \tilde{r}_k)$ is a non-empty set. The definition of $\mathbb{I}_{x_k}(\hat{u}, \hat{\underline{r}}', \tilde{r}_k)$ implies that $(\hat{u}, \hat{\underline{r}}') \in F_k(x_k, \tilde{r}_k)$. Therefore this implies expression (10) holds for any $(u, \underline{r}') \in F_k(x_k, r_k)$. \square

The following Lemma provides a sensitivity condition for the value function $V_k(x_k, r_k)$.

Lemma III.5. *Suppose $F_k(x_k, r_k)$ and $F_k(x_k, \tilde{r}_k)$ are non-empty sets for $k \in \{0, \dots, N-1\}$. Then, for $x_k \in S$, $r_k, \tilde{r}_k \in \Phi_k(x_k)$, such that $r_k \geq \tilde{r}_k$, $k \in \{0, \dots, N\}$, the following expression holds:*

$$0 \leq V_k(x_k, \tilde{r}_k) - V_k(x_k, r_k) \leq M_{V,k}(r_k - \tilde{r}_k) \quad (12)$$

where $M_{V,k} = (M_c + M_q(N - k - 1)c_{\max} + M_{V,k+1})M_{r,k} > 0$, and $M_{V,N} = 0$.

Proof. First, for $k \in \{0, \dots, N-1\}$, when $\tilde{r}_k \leq r_k$, by Lemma IV.1 in [1], we know that $V_k(x_k, \tilde{r}_k) \geq V_k(x_k, r_k)$. The proof is completed if we can show that for $\tilde{r}_k \leq r_k$,

$$V_k(x_k, \tilde{r}_k) - V_k(x_k, r_k) \leq M_{V,k}(r_k - \tilde{r}_k).$$

First, at $k = N$, for any $r_N, \tilde{r}_N \in \Phi_N(x_N)$, we get $V_N(x_N, \tilde{r}_N) = V_N(x_N, r_N) = 0$. Inequality (12) trivially holds for any $M_{V,N} > 0$. By induction's assumption, suppose there exists $M_{V,j+1} > 0$ such that following inequality holds at $k = j + 1$:

$$|V_{j+1}(x, \tilde{r}_{j+1}) - V_{j+1}(x, r_{j+1})| \leq M_{V,j+1} |\tilde{r}_{j+1} - r_{j+1}|.$$

for any $x \in S$. Then, for the case at $k = j$, by Theorem IV.2 in [1], the infimum of $T_j[V_{j+1}]$ is attained. From

² Theorem 4.2 in [16] implies both $\partial P_{x_k, u, \underline{r}'}(r)$, $\partial^\infty P_{x_k, u, \underline{r}'}(r) \subseteq \{0\}$ for $r_k \in \Phi_k(r_k)$. This result further implies $P_{x_k, u, \underline{r}'}(r)$ is strictly differentiable. For details, please refer to this paper.

Theorem II.4, $V_j(x_j, r_j) = T_j[V_{j+1}](x_j, r_j)$. For any given $x_j \in S$, $r_j \in \Phi_j(x_j)$, let $(u_j^*, r^{*,'})$ be the minimizer of $T_j[V_{j+1}](x_j, r_j)$. Then, there exists $(\hat{u}_j, \hat{r}') \in F_j(x_j, \tilde{r}_j)$, such that inequality (10) and the following expressions hold:

$$\begin{aligned} & V_j(x_j, \tilde{r}_j) - V_j(x_j, r_j) \\ & \leq c(x_j, \hat{u}_j) - c(x_j, u_j^*) + \sum_{x' \in S} Q(x'|x_j, \hat{u}_j) V_{j+1}(x', \hat{r}') \\ & \quad - \sum_{x' \in S} Q(x'|x_j, u_j^*) V_{j+1}(x', r^{*,'}) \\ & = c(x_j, \hat{u}_j) - c(x_j, u_j^*) \\ & \quad + \sum_{x' \in S} Q(x'|x_j, \hat{u}_j) (V_{j+1}(x', \hat{r}') - V_{j+1}(x', r^{*,'})) \\ & \quad + \sum_{x' \in S} (Q(x'|x_j, \hat{u}_j) - Q(x'|x_j, u_j^*)) V_{j+1}(x', r^{*,'}) \\ & \leq \|V_{j+1}\|_\infty \sum_{x' \in S} |Q(x'|x_j, \hat{u}_j) - Q(x'|x_j, u_j^*)| \\ & \quad + \sum_{x' \in S} \left\{ |V_{j+1}(x', r^{*,'}) - V_{j+1}(x', \hat{r}')| \right\} \\ & \quad + |c(x_j, \hat{u}_j) - c(x_j, u_j^*)|. \end{aligned}$$

The first inequality follows from the definitions. The second inequality follows from $\sum_{x' \in S} Q(x'|x_j, \hat{u}_j) = 1$ and the definition of $\|V_{j+1}\|_\infty$ and c_{\max} . From Assumption (1) and Inductions' assumption, the above expression further implies

$$\begin{aligned} & V_j(x_j, \tilde{r}_j) - V_j(x_j, r_j) \\ & \leq (M_c + M_q \|V_{j+1}\|_\infty) |\hat{u}_j - u_j^*| \\ & \quad + M_{V,j+1} \sum_{x' \in S} |\hat{r}'(x') - r^{*,'}(x')| \\ & \leq (M_c + M_q \|V_{j+1}\|_\infty + M_{V,j+1}) M_{r,j} |\tilde{r}_j - r_j|. \quad (13) \end{aligned}$$

The last inequality is simply resulted from by Lemma III.4. In addition, from Lemma III.3, we get

$$\|V_{j+1}\|_\infty = \sum_{i=j+1}^{N-1} \|V_i\|_\infty - \|V_{i+1}\|_\infty \leq (N - j - 1)c_{\max}.$$

Then, by applying this inequality to the expression derived in the previous part of the proof, we get

$$\begin{aligned} & V_j(x_j, \tilde{r}_j) - V_j(x_j, r_j) \\ & \leq (M_c + M_q(N - j - 1)c_{\max} + M_{V,j+1}) M_{r,j} |\tilde{r}_j - r_j|. \end{aligned} \quad (14)$$

Thus by induction, expression (12) holds. \square

The next Lemma shows that the difference between dynamic programming operators $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k)$ and $T_k[V_{k+1}](x_k, r_k)$ is bounded.

Lemma III.6. *For any $x_k \in S$, $r_k \in \Phi_k(x_k)$, the following inequality holds for $k \in \{0, \dots, N-1\}$:*

$$0 \leq \bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k) - T_k[V_{k+1}](x_k, r_k) \leq M_{V,k+1} \Delta$$

where $M_{V,k+1} > 0$ is given by Lemma III.5 and Δ is the step size of the discretization of risk threshold r_k .

Proof. First, by the definition of $F_k^D(x_k, r_k)$, we know that $F_k^D(x_k, r_k) \subseteq F_k(x_k, r_k)$. Since, the objective functions and all other constraints in $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k)$ and

$T_k[V_{0,k+1}](x_k, r_k)$ are identical, we can easily conclude that $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k) \geq T_k[V_{k+1}](x_k, r_k)$ for all $x_k \in S$, $r_k \in \Phi_k(x_k)$. The proof is completed if we can show

$$\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k) - T_k[V_{k+1}](x_k, r_k) \leq M_{V,k+1}\Delta.$$

By Theorem IV.2 in [1] we know that the infimum of $T_k[V_{k+1}](x_k, r_k)$ is attained. Let $(u_k^*, r^{*,'}) \in F_k(x_k, r_k)$ be the minimizer of $T_k[V_{k+1}](x_k, r_k)$. Also, for every fixed $x' \in S$, let $\tau(x') \in \{0, \dots, t\}$ such that $r^{*,'}(x') \in \Phi_{k+1}^{(\tau(x'))}(x')$. Now, construct

$$\tilde{r}'(x') := r_{k+1}^{(\tau(x'))} \in \Phi_{k+1}^{(\tau(x'))}(x').$$

By definition of $\bar{\Phi}_{k+1}(x')$, we know that $\tilde{r}'(x') \in \bar{\Phi}_{k+1}(x')$, $\forall x' \in S$. Since $r_{k+1}^{(\tau(x'))}$ is the lower bound of $\Phi_{k+1}^{(\tau(x'))}(x')$, we have $r_{k+1}^{(\tau(x'))} \leq r^{*,'}(x')$. Furthermore, since the size of $\Phi_{k+1}^{(\tau(x'))}(x')$ is Δ , we know that $|r_{k+1}^{(\tau(x'))} - r^{*,'}(x')| \leq \Delta$ for any $x' \in S$. By monotonicity of coherent risk measures,

$$d(x_k, u_k^*) + \rho_k(\tilde{r}'(x_{k+1})) \leq d(x_k, u_k^*) + \rho_k(r^{*,'}(x_{k+1})) \leq r_k.$$

Therefore, we conclude that $(u_k^*, \tilde{r}') \in F_k^D(x_k, r_k)$ is a feasible solution to the problem in $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k)$. From this fact, we get the following inequalities:

$$\begin{aligned} & \bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k) - T_k[V_{k+1}](x_k, r_k) \\ & \leq \sum_{x' \in S} Q(x'|x_k, u_k^*) \left(V_{k+1}(x', \tilde{r}'(x')) - V_{k+1}(x', r^{*,'}(x')) \right) \\ & \leq \sup_{x' \in S} \left\{ |V_{k+1}(x', \tilde{r}'(x')) - V_{k+1}(x', r^{*,'}(x'))| \right\} \\ & \leq M_{V,k+1} \sup_{x' \in S} |\tilde{r}'(x') - r^{*,'}(x')| \leq M_{V,k+1}\Delta. \end{aligned}$$

The first inequality is due to substitutions of the feasible solution of $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k)$ and the optimal solution of $T_k[V_{k+1}](x_k, r_k)$. The second inequality is trivial. The third inequality is a result of Lemma III.5 and the fourth inequality is due to the definition of $\tilde{r}'(x')$, for all $x' \in S$. This completes the proof. \square

The following Lemma is the main result of this section. It characterizes the error bound between the dynamic programming operator $T_k[V_{k+1}](x_k, r_k)$ and $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k)$.

Lemma III.7. *Suppose Assumptions (1) to (2) hold. Then, there exists a constant $M_{V,k} > 0$ such that*

$$\|\bar{T}_{\Delta,k}^D[V_{k+1}] - T_k[V_{k+1}]\|_\infty \leq (M_{V,k} + M_{V,k+1})\Delta \quad (15)$$

where $\bar{T}_{\Delta,k}^D[V_{k+1}](x, r)$ is defined in equation (7), Δ is the step size of the discretization of risk threshold r_k and the expression of $M_{V,k}, M_{V,k+1} > 0$ is given in Lemma III.5, for $k \in \{0, \dots, N-1\}$.

Proof. For any given $x_k \in S$ and $r_k \in \Phi_k(x_k)$, let $\tau \in \{0, \dots, t\}$ such that $r_k \in \Phi_k^{(\tau)}(x_k)$. Then, by the definition of $\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k)$ and Theorem II.4, the following expression holds:

$$\begin{aligned} & |\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k) - T_k[V_{k+1}](x_k, r_k)| \leq |V_k(x_k, r_k^{(\tau)}) - \\ & V_k(x_k, r_k)| + |\bar{T}_{\Delta,k}^D[V_{k+1}](x_k, r_k^{(\tau)}) - T_k[V_{k+1}](x_k, r_k^{(\tau)})|. \end{aligned}$$

Also, by using Lemma III.5 and III.6, the above expression implies that

$$\begin{aligned} & |T_{\Delta,k}^D[V_{k+1}](x_k, r_k) - T_k[V_{k+1}](x_k, r_k)| \\ & \leq M_{V,k+1}\Delta + M_{V,k}|r_k - r_k^{(\tau)}| \leq (M_{V,k} + M_{V,k+1})\Delta. \end{aligned}$$

The last inequality follows from the fact that $r_k \in \Phi_k^{(\tau)}(x_k)$ implies $|r_k^{(\tau)} - r_k| \leq \Delta$, where $r_k^{(\tau)}$ is the lower bound of the discretized region of risk threshold: $\Phi_k^{(\tau)}(x_k)$. By taking supremum of $x_k \in S$ and $r_k \in \Phi_k(x_k)$ on both sides of the resultant inequality, we conclude the inequality given in expression (15). \square

Next, define $M_r = \max_{k \in \{0, \dots, N-1\}} M_{r,k}$. The following Theorem provides an error bound between the value function: $V_k(x_k, r_k)$ and the value function with discretizations: $V_k^D(x_k, r_k)$.

Theorem III.8. *Define $V_k^D(x_k, r_k) := \bar{T}_{\Delta,k}^D[V_{k+1}^D](x_k, r_k)$, $k \in \{0, \dots, N-1\}$ as the value function with discretized risk threshold/update where $V_N^D(x_N, r_N) := V_N(x_N, r_N) = 0$. Suppose Assumptions (1) to (2) hold. Then,*

$$\begin{aligned} \|V_k^D - V_k\|_\infty & \leq 2\Delta \left(\frac{(M_r M_q c_{\max} - M_c(1 - M_r))(1 - M_r^N)}{(1 - M_r)^3} \right. \\ & \left. + \frac{N(N-1)M_r M_q c_{\max}}{2(1 - M_r)} + \frac{N(M_c(1 - M_r) - M_q M_r c_{\max})}{(1 - M_r)^2} \right) \end{aligned}$$

where Δ is the step size of the of risk threshold discretization.

Proof. From Theorem III.7, we know that for $j \in \{k, \dots, N-1\}$, $\|T_{\Delta,j}^D[V_{j+1}] - T_j[V_{j+1}]\|_\infty \leq (M_{V,j} + M_{V,j+1})\Delta$, where Δ is the step size of the discretization of risk threshold r_j . Therefore, we have the following expressions:

$$\begin{aligned} \|V_j^D - V_j\|_\infty & = \|T_{\Delta,j}^D[V_{j+1}^D] - T_j[V_{j+1}]\|_\infty \\ & \leq \|T_{\Delta,j}^D[V_{j+1}^D] - T_{\Delta,j}^D[V_{j+1}]\|_\infty + \|T_{\Delta,j}^D[V_{j+1}] - T_j[V_{j+1}]\|_\infty \\ & \leq \|V_{j+1}^D - V_{j+1}\|_\infty + (M_{V,j} + M_{V,j+1})\Delta. \end{aligned}$$

The first equality is due to Theorem II.4 and the fact that $V_j^D(x_j, r_j) = \bar{T}_{\Delta,j}^D[V_{j+1}^D](x_j, r_j)$. The third inequality is based on the non-expansivity property in Lemma III.1 and the arguments in Theorem III.7. Furthermore,

$$\begin{aligned} \|V_k^D - V_k\|_\infty & = \sum_{j=k}^{N-1} \left(\|V_j^D - V_j\|_\infty - \|V_{j+1}^D - V_{j+1}\|_\infty \right) \\ & \leq \left(\sum_{j=k}^{N-1} M_{V,j} + M_{V,j+1} \right) \Delta \leq 2 \left(\sum_{j=0}^{N-1} M_{V,j} \right) \Delta. \end{aligned}$$

Therefore, the proof is completed by summing the right side of the inequality from 0 to $N-1$ and combining all previous arguments. \square

As the step size $\Delta \rightarrow 0$, for any $x_k \in S$ and $r_k \in \Phi_k(x_k)$, this Theorem implies that $V_k^D(x_k, r_k) \rightarrow V_k(x_k, r_k)$.

Remark III.9. *Unfortunately, similar to all multi-grid discretization approaches discussed in [9], [11], [8], the multi-grid discretization algorithm in this paper also suffers from the curse of dimensionality. Suppose the number of discretized grid used is $|R|$. For each time horizon, the size of*

state space is $|S||R|$. However, the size of the action space is $|A|(|R|)^{|S|}$. Methods such as Branch and bound or rollout algorithms can be applied to find the minimizers in each step to alleviate this issue, if the upper/lower bounds of the value functions are effectively calculated.

IV. NUMERICAL IMPLEMENTATION

Consider an example with 3 states ($x \in \{1, 2, 3\}$), 2 available actions ($u \in \{1, 2\}$) with time horizon $N = 3$. The costs, constraint costs and transition probabilities are given as follows:

$$\begin{bmatrix} c(1,1) & c(1,2) \\ c(2,1) & c(2,2) \\ c(3,1) & c(3,2) \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 5 & 6 \end{bmatrix}, \begin{bmatrix} d(1,1) & d(1,2) \\ d(2,1) & d(2,2) \\ d(3,1) & d(3,2) \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 5 & 4 \\ 6 & 3 \\ 5 & 1 \end{bmatrix},$$

$$Q(x'|x, 1) = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}, Q(x'|x, 2) = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}.$$

For any $x_0 \in S$ and $r_0 \in \Phi_0(x_0)$, the risk sensitive constrained stochastic optimal control problem we are solving is as follows:

$$\min_{\pi \in \Pi} \mathbb{E} \left[\sum_{k=0}^2 c(x_k, u_k) \right]$$

$$\text{subject to } \rho_{0,3}(d(x_0, u_0), d(x_1, u_1), d(x_2, u_2), 0) \leq r_0.$$

where $u_k = \pi_k(h_{0,k})$ for $k \in \{0, 1, 2\}$, $\rho_{0,N}(Z_0, Z_1, Z_2, Z_3) = Z_0 + \rho_0(Z_1 + \rho_1(Z_2 + \rho_2(Z_3)))$ and

$$\rho_k(V) = \mathbb{E}[V] + 0.2 \left(\mathbb{E} [[V - \mathbb{E}[V]]_+^2] \right)^{1/2}.$$

First, this problem can be re-casted using multi-stage constrained dynamic programming using the methods described by Theorem IV.3 in [1]. Furthermore, based on equations (7) to (8), we can approximate the optimal value function using risk threshold/update discretization. In this example, we discretize every risk threshold sets into M regions, where

$$M \in \{5, 10, 20, 40, 60, 80, 100, 150\}.$$

With different sizes of risk threshold discretization, we get approximations of optimal value functions, up to various degrees of accuracies. Figure 1 shows both the approximations of value function using various step sizes and their errors of approximations. As the number of M increases, the approximated value function converges towards the true optimal value function. However, as discussed in Remark III.9, the size of action space increases exponentially with the number of states, thus it makes enumerating all state/action pairs during value iteration computationally expensive.

V. CONCLUSION

In this paper we have presented and analyzed a uniform grid discretization algorithm for approximating the Bellman's recursion for finite horizon constrained stochastic optimal control problems. Although the current algorithm suffers from curse of dimensionality, it is by far the only known algorithm for numerically approximating constrained dynamic programming algorithms with continuous risk updates. This paper also leaves important extensions open for further researches that involve randomized grid sampling and variable resolution of discretization.

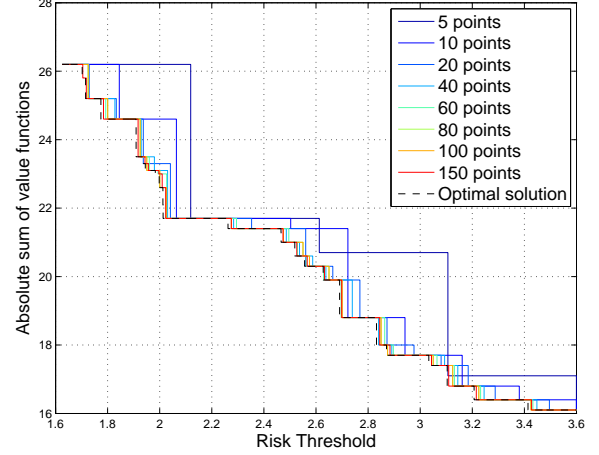


Fig. 1. Convergence of approximated value functions, and errors of approximations.

REFERENCES

- [1] Y. Chow and M. Pavone. Stochastic optimal control with dynamic, time-consistent risk constraints. In *American Control Conference*, 2013.
- [2] D. Bertsekas. *Dynamic programming and optimal control*. Athena Scientific, 2005.
- [3] R. Chen and E. Feinberg. Non-randomized policies for constrained Markov decision process. *Mathematical Methods in Operations Research*, 66:165–179, 2007.
- [4] A. Ruszczyński. Risk-averse dynamic programming for Markov decision process. *Journal of Mathematical Programming*, 125(2):235–261, 2010.
- [5] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.
- [6] P. Huang, D. Iancu, M. Petrik, and D. Subramanian. The price of dynamic inconsistency for distortion risk measures. *arXiv preprint arXiv:1106.6102*, 2011.
- [7] W. Whitt. Approximations of dynamic programs, Volume I. *Mathematics of Operations Research*, 3(3):231–243, 1978.
- [8] G. Gordon. Approximate solutions to Markov decision processes. *Robotics Institute*, page 228, 1999.
- [9] C. Chow and J. Tsitsiklis. An optimal one-way multigrid algorithm for discrete-time stochastic control. *IEEE Transactions on Automatic Control*, 36(8):898–914, 1991.
- [10] J. Rust. Using randomization to break the curse of dimensionality. *Econometrica: Journal of the Econometric Society*, pages 487–516, 1997.
- [11] J. Tsitsiklis and B. Van Roy. An analysis of temporal-difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42(5):674–690, 1997.
- [12] R. Munos and A. Moore. Rates of convergence for variable resolution schemes in optimal control. 2000. Available at <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.1.1.1.1.1.1>.
- [13] R. Munos and A. Moore. Barycentric interpolators for continuous space and time reinforcement learning. In *Neural Information Processing Systems*, 1998.
- [14] R. Munos. A convergent reinforcement learning algorithm in the continuous case: the finite-element reinforcement learning. 1996. Available at http://www.ri.cmu.edu/pub_files/pub1/munos_remi_1996_1/munos_remi_1996_1.pdf.
- [15] N. Nguyen and G. Lafferriere. Lipschitz properties of non-smooth functions and set-valued mappings via generalized differentiation and applications. *arXiv preprint arXiv:1302.1794*, 2013.
- [16] Y. Lucet and J. Ye. Sensitivity analysis for the value function for optimization problems with variational inequalities constraints. *SIAM J. OPTIM*, 40(3):699–723, 2002.
- [17] R. Rockafellar and R. Wets. *Variational Analysis*. Springer, 1998.

